

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2004, 19:24:00 ; Search time 80.25 Seconds
(without alignments)
43.019 Million cell updates/sec

Title: US-10-618-644-3
Perfect score: 38
Sequence: 1 NWGPLV 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues
Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 02:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	257	Q6MP42	Q6mp42 bdellovibri
2	38	100.0	257	Cae78956	Cae78956 bdellovib
3	38	100.0	562	GLC4_SOYBN	P02858 glycine max
4	38	100.0	562	Q43452	Q43452 glycine max
5	38	100.0	563	Q39921	Q39921 glycine soj
6	38	100.0	563	Q9SB11	Q9sb11 glycine max
7	37	97.4	483	Q6CLB6	Q6clb6 kluyveromyc
8	37	97.4	683	Q89259	Q89259 bacteroides
9	36	94.7	724	Q788R9	Q788r9 neurospora
10	36	94.7	793	Q6F9M7	Q6f9m7 acinetobact
11	35	92.1	204	CEA_ECOLI	Q47108 escherichia
12	35	92.1	378	Q7ZA77	Q7za77 penicillium
13	35	92.1	387	CEAN_ECOLI	P08083 escherichia
14	35	92.1	499	Q9XB77	Q9xb77 escherichia
15	35	92.1	510	CEAB_ECOLI	P05819 escherichia
16	35	92.1	592	CEA_CITR	P04480 citrobacter
17	35	92.1	611	Q8SBG9	Q8sb9g bacterioph
18	35	92.1	619	Q24681	Q24681 shigella bo
19	35	92.1	629	Q9KJ98	Q9kj98 escherichia
20	35	92.1	657	Q8JUN0	Q8jnm0 aspergillus
21	35	92.1	666	Q7SDR8	Q7sdr8 neurospora
22	35	92.1	858	Q7Q6R9	Q7q6r9 anopheles g
23	35	92.1	1073	Q8GUN3	Q8gun3 caenorhabdi
24	34	89.5	178	Q6NDI7	Q6ndi7 rhodopsu
25	34	89.5	178	Q8AE25565	Cae25565 rhodopseu
26	34	89.5	250	Q7DM64	Q7dm64 arabisdopeis
27	34	89.5	261	Q6HMR9	Q6hmr9 bacillus th
28	34	89.5	265	Q9K7I7	Q9k7i7 bacillus ha
29	34	89.5	325	Q9AM28	Q9am28 xanthomonas
30	34	89.5	336	Q9PET8	Q9pet8 xanthomonas
31	34	89.5	340	O04348	O04348 arabisdopeis

32	34	89.5	344	1	HIFN_BRARE	P59723 brachydanio
33	34	89.5	367	2	Q8LEC9	Q8lec9 arabidopsis
34	34	89.5	367	2	Q9M9Z2	Q9m9z2 arabidopsis
35	34	89.5	431	2	Q8ZT65	Q8zt65 pyrobaculum
36	34	89.5	449	2	Q7PT63	Q7pt63 anopheles g
37	34	89.5	464	2	Q6EQK9	Q6eqk9 oryza sativ
38	34	89.5	470	2	Q05351	Q05351 entamoeba h
39	34	89.5	470	2	Q7Y0D0	Q7y0d0 oryza sativ
40	34	89.5	499	2	Q7UY60	Q7uy60 rhodopirell
41	34	89.5	539	2	Q9USK3	Q9usk3 schizosacch
42	34	89.5	586	2	Q67335	O67335 aquifex aeo
43	34	89.5	593	2	Q66826	O66826 aquifex aeo
44	34	89.5	616	2	Q7KR41	Q7kr41 drosophila
45	34	89.5	616	2	AAW68656	AAW68656 drosophila

ALIGNMENTS

RESULT 1
ID Q6MP42 PRELIMINARY; PRT; 257 AA.
AC Q6MP42;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE ABC-type transport system permease protein.
GN Name=glpF; OrderedLocusNames=Bdl1024;
OS Bdellovibrio bacteriovorus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;
OC Bdellovibrionaceae; Bdellovibrio.
OX NCBI_TaxID=959;
RN [1]
RP -SEQUENCE FROM N.A.
RC STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;
RX PubMed=14752164;
RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,
RA Kellner H., Lambert C., Evans K.J., Goesmann A., Meyer F.,
RA Sockett R.E., Schuster S.C.;
RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a
genomic perspective".
RL Science 303:689-692(2004).
DR EMBL; BX842648; Cae78956.1; -.
KW Complete proteome.
SQ SEQUENCE 257 AA; 28762 MW; 3006017ED9B2F0E7 CRC64;

Query Match 100.0%; Score 38; DB 2; Length 257;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NWGPLV 6
|||
Db 146 NWGPLV 151

RESULT 2
ID Cae78956 PRELIMINARY; PRT; 257 AA.
AC Cae78956;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DE ABC-type transport system permease protein.
GN GLDF OR Bdl024.
OS Bdellovibrio bacteriovorus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;
OC Bdellovibrionaceae; Bdellovibrio.
OX NCBI_TaxID=959;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;
RX PubMed=14752164;
RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,

RA Keller H., Lambert C., Evans K.J., Goessmann A., Meyer F.,
 RA Sockett R.E., Schuster S.C.;
 RT "A predator unmasked: life cycle of *Bdellovibrio bacteriovorus* from a
 RT genomic perspective.";
 RL Science 303:689-692(2004).
 DR EMBL; BX842648; CAB78956.1; -.
 SQ SEQUENCE 257 AA; 28762 MW; 3006017ED9B2F0E7 CRC64;

Query Match 100.0%; Score 38; DB 2; Length 257;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NWGPLV 6
 |||||
 Db 146 NWGPLV 151

RESULT 3
 ID GLC4_SOYBN STANDARD; PRT; 562 AA.
 AC P02858
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Glycinin G4 precursor [Contains: Glycinin A5 subunit; Glycinin A4
 DE subunit; Glycinin B3 subunit].
 GN Name-GY4;
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 OX NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Bonninori;
 RA Hirano H., Fukazawa C., Harada K.;
 RT "The primary structures of the A4 and A5 subunits are highly
 RT homologous to that of the A3 subunit in the glycinin seed storage
 RT protein of soybean.";
 RL FEBS Lett. 181:124-128(1985).
 RN [3]
 RP SEQUENCE (A4/A5 SUBUNITS).
 RC STRAIN=cv. Bonninori;
 RA Hirano H., Fukazawa C., Harada K.;
 RT "The primary structures of the A4 and A5 subunits are highly
 RT homologous to that of the A3 subunit in the glycinin seed storage
 RT protein of soybean.";
 RL FEBS Lett. 181:124-128(1985).
 RN [3]
 RP SEQUENCE OF 181-386 FROM N.A.
 RC STRAIN=cv. CX635-1-1-1;
 RA Scallion B.J., Dickinson C.D., Nielsen N.C.;
 RT "Characterization of a null-allele for the G₄ glycinin gene from
 RT soybean.";
 RL Mol. Gen. Genet. 208:107-113(1987).
 CC -!- FUNCTION: Glycinin is the major seed storage protein of soybean.
 CC -!- SUBUNIT: Hexamer; each subunit is composed of an acidic and a
 CC basic chain derived from a single precursor and linked by a
 CC disulfide bond.
 CC -!- SIMILARITY: Belongs to the 11S seed storage protein (globulins)
 CC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; X02626; CAA326478.1; -.
 EMBL; X05652; CAB57802.1; -.
 PIR; A91145; FWSYG5.

DR HSP; P04776; 1FXZ.
 DR InterPro; IPR006045; Cupin.
 DR InterPro; IPR011051; RmlC_like_cupin.
 DR InterPro; IPR006044; Seedstore_118.
 DR Pfam; PF00190; Cupin; 2.
 DR PRINTS; PR00439; 11SGLOBULIN.
 DR PROSITE; PS00305; 11S SEED STORAGE; 1.
 KW Direct protein sequencing; Multigene family; Seed storage protein;
 .Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 120
 FT CHAIN 121 377
 FT CHAIN 378 562
 FT CHAIN 562 562
 FT DISULFID 108 384
 FT CONFLICT 29 29
 FT CONFLICT 82 82
 FT CONFLICT 86 86
 FT CONFLICT 94 94
 FT CONFLICT 101 103
 FT CONFLICT 105 105
 FT CONFLICT 117 117
 FT CONFLICT 253 253
 FT CONFLICT 332 335
 FT CONFLICT 332 335
 SQ SEQUENCE 562 AA; 63587 MW; F5A06B8856B9BBD6 CRC64;

Query Match 100.0%; Score 38; DB 1; Length 562;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NWGPLV 6
 |||||
 Db 543 NWGPLV 548

RESULT 4
 Q43452 PRELIMINARY; PRT; 562 AA.
 AC Q43452;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Glycinin precursor.
 GN Name=Gy4;
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 OX NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RX MEDLINE=92256811; PubMed=1316192;
 RA Xue Z.T., Xu M.L., Shen W., Zhuang N.L., Hu W.M., Shen S.C.;
 RT "Characterization of the G₄ Glycine gene from soybean Glycine max cv.
 RT Forrest.";
 RL Plant Mol. Biol. 18:897-908(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RA Hue Z.T.;
 RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBUNIT: Hexamer; each subunit is composed of an acidic and a
 CC basic chain derived from a single precursor and linked by a
 CC disulfide bond (By similarity).
 CC -!- SIMILARITY: Belongs to the 11S seed storage protein (globulins)
 CC family.
 DR EMBL; X52863; CAA37044.1; -.
 DR PIR; PQ0199; PQ0199.
 DR PIR; S20946; S20946.
 DR HSP; P04776; 1FXZ.
 DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
 DR InterPro; IPR006045; Cupin.
 DR InterPro; IPR011051; RmlC_like_cupin.

```

DR InterPro; IPR006044; Seedstore_11s.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S_SEED_STORAGE; 1.
KW Signal.
FT CHAIN 1 23 Potential.
FT CHAIN 24 562 glycinin.
SQ SEQUENCE 562 AA; 63876 MW; 3A4EF28E448AB15A CRC64;

Query Match 100.0%; Score 38; DB 2; Length 562;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NMGPLV 6
DB 543 NMGPLV 548
|||||

RESULT 5
Q39921 PRELIMINARY; PRT; 563 AA.
AC Q39921;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE A5A4B3 subunit.
OS Name=glycinin G4;
GN Glycine soja (Wild soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3848;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SH1;
RA Xue Z.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: Hexamer; each subunit is composed of an acidic and a
CC basic chain derived from a single precursor and linked by a
CC disulfide bond (By similarity).
CC -1- SIMILARITY: Belongs to the 11S seed storage protein (globulins)
CC family.
EMBL; X86970; CAA60533.1; -.
DR PIR; S54802; S54802.
DR HSP; P04776; 1PXZ.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR011051; RmlC like cupin.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S_SEED_STORAGE; 1.
DR SEQUENCE 563 AA; 63806 MW; EC171BED5F2B361F CRC64;

Query Match 100.0%; Score 38; DB 2; Length 563;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NMGPLV 6
DB 544 NMGPLV 549
|||||

RESULT 6
Q9SB11 PRELIMINARY; PRT; 563 AA.
AC Q9SB11;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycinin.
OS Glycine max (soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

DR InterPro; IPR006044; Seedstore_11s.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S_SEED_STORAGE; 1.
DR SEQUENCE 562 AA; 63876 MW; 3A4EF28E448AB15A CRC64;

Query Match 100.0%; Score 38; DB 2; Length 562;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NMGPLV 6
DB 543 NMGPLV 548
|||||

RESULT 5
Q39921 PRELIMINARY; PRT; 563 AA.
AC Q39921;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE A5A4B3 subunit.
OS Name=glycinin G4;
GN Glycine soja (Wild soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3848;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SH1;
RA Xue Z.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: Hexamer; each subunit is composed of an acidic and a
CC basic chain derived from a single precursor and linked by a
CC disulfide bond (By similarity).
CC -1- SIMILARITY: Belongs to the 11S seed storage protein (globulins)
CC family.
EMBL; X86970; CAA60533.1; -.
DR PIR; S54802; S54802.
DR HSP; P04776; 1PXZ.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR011051; RmlC like cupin.
DR InterPro; IPR006044; Seedstore_11s.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S_SEED_STORAGE; 1.
DR SEQUENCE 563 AA; 63806 MW; EC171BED5F2B361F CRC64;

Query Match 100.0%; Score 38; DB 2; Length 563;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NMGPLV 6
DB 544 NMGPLV 549
|||||

RESULT 6
Q9SB11 PRELIMINARY; PRT; 563 AA.
AC Q9SB11;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycinin.
OS Glycine max (soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

DR InterPro; IPR006044; Seedstore_11s.
DR euroids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaves;
RC Chen S., Arahira M., Fukazawa C.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: Hexamer; each subunit is composed of an acidic and a
CC basic chain derived from a single precursor and linked by a
CC disulfide bond (By similarity).
CC -1- SIMILARITY: Belongs to the 11S seed storage protein (globulins)
CC family.
EMBL; AB004062; BAA74953.1; -.
DR PIR; PQ0199; PQ0199.
DR HSP; P04776; 1PXZ.
DR GO; GO:0045735; F:nutrient reservoir activity; IEA.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR011051; RmlC like cupin.
DR InterPro; IPR006044; Seedstore_11s.
DR Pfam; PF00190; Cupin; 2.
DR PRINTS; PR00439; 11SGLOBULIN.
DR PROSITE; PS00305; 11S_SEED_STORAGE; 1.
DR SEQUENCE 563 AA; 63797 MW; 4ACC765C55AB9E18 CRC64;

Query Match 100.0%; Score 38; DB 2; Length 563;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NMGPLV 6
DB 544 NMGPLV 549
|||||

RESULT 7
Q6CLB6 PRELIMINARY; PRT; 483 AA.
AC Q6CLB6;
DT 01-OCT-2004 (TrEMBLrel. 28, Created)
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to ca|CA6037|CaERC4 Candida albicans ethionine resistance
DE protein.
DE ORFNames=KLLA0F04279g;
GN Kluyveromyces lactis (Yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RG GENOLEVURES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boissame A., Boyer E., Cattolico L., Confaniolieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikoleki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-1140;
RA Genoscope;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.

```

DR EMBL: CR382126; CAG97981.1; -. B754A50ECA8CB158 CRC64;
SQ SEQUENCE 483 AA; 52885 MW; 754A50ECA8CB158 CRC64;
Query Match 97.4%; Score 37; DB 2; Length 483;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 NWGPIV 6
Db 261 NWGPII 266
RESULT 8
Q89ZS9 PRELIMINARY; PRT; 683 AA.
AC Q89ZS9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative outer membrane protein, probably involved in nutrient binding.
GN OrderedLocNames=B74297;
OS Bacteroides thetaiotaomicron.
OC Bacteri; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."
RL Science 299:2074-2076(2003).
DR EMBL; AE016944; AA079402.1; -.
KW Complete proteome.
SQ SEQUENCE 683 AA; 77762 MW; D2541E243339AF88 CRC64;
Query Match 97.4%; Score 37; DB 2; Length 683;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 NWGPIV 6
Db 158 NWGPII 163
RESULT 9
Q7S8R9 PRELIMINARY; PRT; 724 AA.
ID Q7S8R9
AC Q7S8R9;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU07720.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OK74R;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehm B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysysselis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Ramussen C., Metznerberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,

RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa."
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000225; EAA32734.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0015297; F:antipporter activity; IEA.
DR GO; GO:0015238; F:drug transporter activity; IEA.
DR GO; GO:0006855; P:multidrug transport; IEA.
DR InterPro; IPR002528; MatE.
DR Pfam; PF01554; MatE; 2.
DR TIGRPFAMs; TIGR00797; matE; 1.
KW Hypothetical protein.
SQ SEQUENCE 724 AA; 78571 MW; 643E753D68BD47F8 CRC64;
Query Match 94.7%; Score 36; DB 2; Length 724;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 NWGPIV 6
Db 444 NWGPIV 449
RESULT 10
Q6F9M7 PRELIMINARY; PRT; 793 AA.
ID Q6F9M7
AC Q6F9M7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Xanthine dehydrogenase, large subunit (EC 1.1.1.204).
GN Name=xdhB; OrderedLocNames=ACIAD2467;
OS Acinetobacter sp. (strain ADP1).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_TaxID=62977;
RN [1]
RP SEQUENCE FROM N.A.
RA Barbe V., Vallenet D., Fonknechten N., Kreimeyer A., Oztas S.,
RA Labarre L., Cruveiller S., Robert C., Duprat S., Wincker P.,
RA Ornston L.N., Weissenbach J., Marliere P., Cohen G.N., Medigue C.;
RT "Unique features revealed by the genome sequence of Acinetobacter sp.
ADP1, a versatile and naturally transformation competent bacterium."
RL Nucleic Acids Res. 0:0-0(2004).
DR EMBL; CR543861; CAG69237.1; -.
DR InterPro; IPR008274; Aldxan_dh_bind.
DR InterPro; IPR000674; Aldxan_dh_hamm.
DR Pfam; PF01315; Aldxan_dh_C; 1.
DR Pfam; PF02738; Aldxan_dh_C2; 1.
DR Complete proteome; Oxidoreductase.
SQ SEQUENCE 793 AA; 87151 MW; E577A75B87D16D86 CRC64;
Query Match 94.7%; Score 36; DB 2; Length 793;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 NWGPIV 6
Db 89 NWGPIV 94
RESULT 11
CEA_ECOLI STANDARD; PRT; 204 AA.
ID CEA_ECOLI
AC Q47108;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)

DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Colicin A (Fragment).
GN Name=caa;
OS Escherichia coli.
OG Plasmid pColA9.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX MEDLINE=87147231; PubMed=2434951;
RA Baty D., Knibiehler M., Verheij H., Pattus F., Shire D., Bernadac A.,
RA Lazdunski C.;
RT "Site-directed mutagenesis of the COOH-terminal region of colicin A:
RT effect on secretion and voltage-dependent channel activity.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:1152-1156(1987).
RN [2]
RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=92235820; PubMed=1373773;
RA Parker M.W., Poetma J.P.M., Pattus F., Tucker A.D., Tsernoglou D.;
RT "Refined structure of the pore-forming domain of colicin A at 2.4-A
RT resolution.";
RL J. Mol. Biol. 224:639-657(1992).
RN [3]
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M15691; AAA23592.1; --
DR PIR; I53544; I41169.
DR PDB; 1COL; X-ray; A/B=1-204.
DR InterPro; IPR000293; Channel_colicin.
DR Pfam; PF01024; Colicin; 1.
DR PRINTS; PR00280; CHANCOLICIN
DR ProDom; PD002657; Channel_colicin; 1.
DR PROSITE; PS00276; CHANNEL_COLICIN; 1.
KW 3D-structure; Antibiotic; Bacteriocin; Plasmid; Transmembrane.
FT NON TER 1
FT TRANSMEM 139 161 Potential.
FT TRANSMEM 165 187 Potential.
FT SEQUENCE 204 AA; 21790 MW; F7702455E2E4E9B9 CRC64;
Query Match 92.1%; Score 35; DB 1; Length 204;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 NWGPLV 6
Db 129 NWGPLM 134
RESULT 12
ID Q7ZA77 PRELIMINARY; PRT; 378 AA.
AC Q7ZA77;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Exo-arabinanase.
GN Name=abnx;
OS Penicillium chrysogenum (Penicillium notatum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium;

OC Penicillium chrysogenum complex.
OX NCBI_TaxID=5076;
RN [1]
RN SEQUENCE FROM N.A.
RA Inara H., Shibano A., Sakamoto T.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RA Hideshi I., Asako S., Tateuji S.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB096108; BAC76689.1; --
DR InterPro; IPR002860; Glyco_hydro_BNR.
DR Pfam; PF02012; BNR; 2.
SQ SEQUENCE 378 AA; 41694 MW; C780B1E1F8BC8D2A CRC64;
Query Match 92.1%; Score 35; DB 2; Length 378;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 NWGPLV 6
Db 211 NWGPPV 216
RESULT 13
CEAN_ECOLI STANDARD; PRT; 387 AA.
ID P08083;
AC P08083;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Colicin N.
DE Namescna;
GN Escherichia coli.
OG Plasmid COLN pCHAP4.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=88201670; PubMed=2834623;
RA Pugsley A.P.;
RT "Nucleotide sequencing of the structural gene for colicin N reveals
RT homology between the catalytic, C-terminal domains of colicins A and
RT N.";
RL Mol. Microbiol. 1:317-325(1987).
RN [2]
RN SEQUENCE OF 372-387 FROM N.A.
RX STRAIN=K12;
RX MEDLINE=88174431; PubMed=3280946;
RA Pugsley A.P.;
RT "The immunity and lysis genes of ColN plasmid pCHAP4.";
RL Mol. Gen. Genet. 211:335-341(1988).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 91-387.
RX MEDLINE=98362590; PubMed=9687368;
RA Vetter I.R., Parker M.W., Tucker A.D., Lakey J.H., Pattus F.;
RA Tsernoglou D.;
RT "Crystal structure of a colicin N fragment suggests a model for
RT toxicity.";
RL Structure 6:863-874(1998).
CC -1- FUNCTION: This colicin is a channel-forming colicin. This class of
CC transmembrane toxins depolarize the cytoplasmic membrane, leading
CC to dissipation of cellular energy.
CC -1- FUNCTION: Colicins are polypeptide toxins produced by and active
CC against, Escherichia coli and closely related bacteria.
CC -1- SIMILARITY: Belongs to the channel forming colicin family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

-----
DR EMBL; Y00533; CAA68592.1; -.
DR EMBL; X06933; CAA30021.1; -.
DR PIR; S00867; S00867.
DR PDB; 1A87; X-ray; @=67-387.
DR InterPro; IPR000293; Channel_colicin.
DR Pfam; PF01024; Colicin; 1.
DR PRINTS; PR00280; CHANLCOLICIN.
DR ProDom; PD002657; Channel_colicin; 1.
DR PROSITE; PS00276; CHANNEL_COLICIN; 1.
KW 3D-structure; Antibiotic; Bacteriocin; Plasmid; Transmembrane.
FT DOMAIN 1 49
FT TRANSMEM 325 345 Gly-rich.
FT TRANSMEM 350 370 Potential.
FT STRAND 93 94 Potential.
FT TURN 95 96
FT TURN 97 101
FT STRAND 103 104
FT TURN 106 107
FT STRAND 108 111
FT HELIX 114 116
FT TURN 117 119
FT STRAND 122 128
FT TURN 129 130
FT STRAND 131 137
FT TURN 140 141
FT STRAND 143 148
FT HELIX 149 151
FT HELIX 153 155
FT STRAND 157 161
FT HELIX 170 214
FT HELIX 216 230
FT TURN 231 231
FT HELIX 234 236
FT HELIX 240 251
FT TURN 252 252
FT TURN 254 255
FT HELIX 260 271
FT TURN 272 272
FT HELIX 275 283
FT TURN 284 285
FT HELIX 287 289
FT TURN 292 293
FT TURN 294 309
FT HELIX 310 311
FT HELIX 315 326
FT TURN 327 328
FT HELIX 331 343
FT TURN 344 344
FT TURN 348 349
FT HELIX 352 370
FT HELIX 372 385
SQ SEQUENCE 387 AA; 41743 MW; 1C4342B222F8CECD CRC64;

```

Query Match 92.1%; Score 35; DB 1; Length 387;
 Best Local Similarity 83.3%; Pred. No. 2.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NWGPLY 6
 Db 313 NWGPLL 318

```

RESULT 14
Q9XB47 PRELIMINARY; PRT; 499 AA.
AC Q9XB47;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

```

```

DE Colicin S4.
GN Names=csa;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9287842; PubMed=10348872;
RA Pilsel H., Maje D., Braun V.;
RT "Characterization of colicin S4 and its receptor OmpW, a minor protein
  of the Escherichia coli outer membrane.";
RL J. Bacteriol. 181:3578-3581(1999).
DR EMBL; Y18684; CAB46008.1; -.
DR HSSP; P04480; 1COL.
DR InterPro; IPR000293; Channel_colicin.
DR InterPro; IPR001232; Skpl.
DR Pfam; PF01024; Colicin; 1.
DR PRINTS; PR00280; CHANLCOLICIN.
DR ProDom; PD002657; Channel_colicin; 1.
DR PROSITE; PS00276; CHANNEL_COLICIN; 1.
SQ SEQUENCE 499 AA; 54085 MW; 3E36C7271BF1D293 CRC64;

Query Match 92.1%; Score 35; DB 2; Length 499;
Best Local Similarity 83.3%; Pred. No. 3.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NWGPLY 6
Db 424 NWGPLM 429

RESULT 15
CEAB_ECOLI STANDARD; PRT; 510 AA.
ID CEAB_ECOLI
AC P05819;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Colicin B.
OS Name=cba;
OS Escherichia coli.
OG Plasmid ColBM-pF166.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=87250309; PubMed=2439491;
RA Schramm E., Mende J., Braun V., Kamp R.M.;
RT "Nucleotide sequence of the colicin B activity gene cba: consensus
  pentapeptide among TonB-dependent colicins and receptors.";
RL J. Bacteriol. 169:3350-3357(1987).
CC -!- FUNCTION: This colicin is a channel-forming colicin. This class of
  transmembrane toxins depolarize the cytoplasmic membrane, leading
  to dissipation of cellular energy.
CC -!- FUNCTION: Colicins are polypeptide toxins produced by and active
  against Escherichia coli and closely related bacteria.
CC -!- MISCELLANEOUS: This colicin requires tonB for its uptake.
CC -!- SIMILARITY: Belongs to the channel forming colicin family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

-----
DR EMBL; M16816; AAA98063.1; -.
DR PIR; A27089; IKECBB.
DR HSSP; P04480; 1COL.
DR InterPro; IPR000293; Channel_colicin.

```

DR InterPro; IPR003058; Cloacin.
DR InterPro; IPR010916; TonB_Box_N.
DR InterPro; IPR010917; TonB_recept_C.
DR Pfam; PF03515; Cloacin; 1.
DR Pfam; PF01024; Colicin; 1.
DR PRINTS; PR00280; CHANCOLICIN.
DR PRINTS; PR01295; CLOACIN.
DR ProDom; PD002657; Channel_colicin; 1.
DR PROSITE; PS00276; CHANNEL_COLICIN; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW Antibiotic; Bacteriocin; Direct protein sequencing; Plasmid; TonB box;
KW Transmembrane.
FT INIT_MET 0 0
FT SITE_16 23 TonB box.
FT TRANSMEM 454 474 Potential.
FT TRANSMEM 476 496 Potential.
SQ SEQUENCE 510 AA; 54732 MW; 6E4B972CF19245F1 CRC64;

Query Match 92.1%; Score 35; DB 1; Length 510;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NWGPLV 6
| | | | |
Db 438 NWGPLM 443

Search completed: November 6, 2004, 19:52:51
Job time : 83.25 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)